Weight Loss

Charlie's Angels

2024-05-29

```
knitr::opts_chunk$set(warning = FALSE, message = FALSE)
```

A. Gym as a factor in Weight Loss

Question #1: On the average, which Gym has individuals with higher weight loss?

```
library(readr)
library(dplyr)
library(rstatix)
library(tinytex)
library(ggplot2)

typediet <- c("A", "B", "C")
typegym <- c("Pewter", "Cerulean")


WEIGHTLOSS <- read_csv("WeightLoss.csv",
                       col_types = cols(Diet = col_factor(levels = typediet),
                                        Gym = col_factor(levels = typegym)
                                       )
                      )
head(WEIGHTLOSS)
```

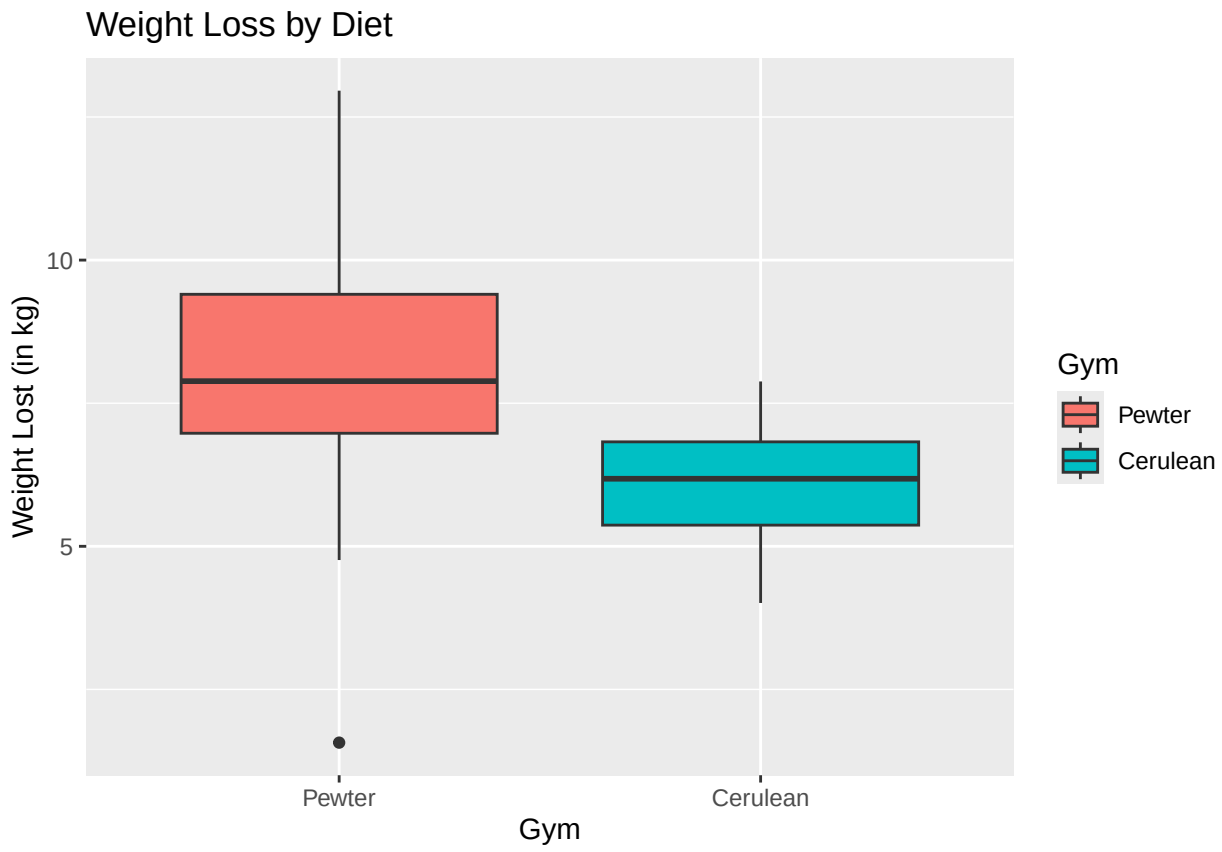
```
## # A tibble: 6 x 5
##   MemberID   Age Diet  Gym      WeightLoss
##      <dbl> <dbl> <fct> <fct>         <dbl>
## 1        1    35 A     Cerulean       5.74
## 2        2    29 C     Cerulean       7.36
## 3        3    27 B     Pewter         7.17
## 4        4    23 C     Pewter        11.9
## 5        5    26 B     Pewter         8.78
## 6        6    32 B     Cerulean       5.52
```

```
mean_gym <- WEIGHTLOSS %>%
  group_by(Gym) %>%
  summarize(mean_value = mean(WeightLoss, na.rm = TRUE))
mean_gym
```

```
## # A tibble: 2 x 2
```

```
##   Gym       mean_value
##   <fct>          <dbl>
## 1 Pewter          8.01
## 2 Cerulean        6.04
```

```
ggplot(WEIGHTLOSS, aes(x = Gym, y = WeightLoss, fill = Gym)) +
  geom_boxplot() +
  labs(title = "Weight Loss by Diet", x = "Gym", y = "Weight Lost (in kg)")
```



Answer: On average, Pewter gym has individuals with higher weight loss

Question #2: At 0.05 level of significance, is there a difference in the average weight loss between the members of the two Gyms? Test for assumptions before performing the T test for means.

```
library(rstatix)
library(tidyverse)

WL2 <- subset(WEIGHTLOSS, Gym %in% c("Pewter", "Cerulean"))|>
  select(WeightLoss, Gym)|>
  mutate(Gym = as.factor(Gym))
WL2
```

```
## # A tibble: 60 x 2
##    WeightLoss Gym
##         <dbl> <fct>
##  1       5.74 Cerulean
```

```
##  2       7.36 Cerulean
##  3       7.17 Pewter
##  4      11.9  Pewter
##  5       8.78 Pewter
##  6       5.52 Cerulean
##  7      10.2  Pewter
##  8       7.77 Cerulean
##  9       7.79 Pewter
## 10       5.49 Cerulean
## # i 50 more rows
```

```
shapiro.test(WL2$WeightLoss)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  WL2$WeightLoss
## W = 0.97623, p-value = 0.2902
```

```
summary_gym <- WL2%>%
  group_by(Gym)%>%
  summarize(shapiro_test(WeightLoss))
summary_gym
```

```
## # A tibble: 2 x 4
##   Gym      variable   statistic p.value
##   <fct>    <chr>          <dbl>   <dbl>
## 1 Pewter   WeightLoss     0.967   0.468
## 2 Cerulean WeightLoss     0.958   0.282
```

The weight loss data from the 2 gyms follows a normal distribution since the p-values of the 2 gyms are greater than 0.05

```
t.test(WeightLoss ~ Gym, WL2,
       var.equal = F,
       alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  WeightLoss by Gym
## t = 4.333, df = 43.356, p-value = 8.58e-05
## alternative hypothesis: true difference in means between group Pewter and group Cerulean is not equal
## 95 percent confidence interval:
##  1.050836 2.879830
## sample estimates:
##   mean in group Pewter mean in group Cerulean
##               8.009333               6.044000
```

Answer: At 0.05 level of significance, we have sufficient evidence to conclude that the average weight loss between the members of the 2 gyms are not equal

B. Diet as a factor in Weight Loss

Question #3: Obtain mean of WeightLoss per Diet. Which types of diet have greater mean weight loss than the overall mean weight loss? Which types of diet have less mean weight loss than the overall mean weight loss?

```
mean_diet <- WEIGHTLOSS %>%
  group_by(Diet) %>%
  summarize(mean_diet = mean(WeightLoss))
mean_diet
```

```
## # A tibble: 3 x 2
##   Diet  mean_diet
##   <fct>     <dbl>
## 1 A          5.82
## 2 B          7.05
## 3 C          8.21
```

```
OVmean <- WEIGHTLOSS %>%
          summarize(OVmean = mean(WeightLoss))
OVmean
```

```
## # A tibble: 1 x 1
##   OVmean
##    <dbl>
## 1   7.03
```

Answer: Among the three diets, Diet B & C has greater mean than the overall mean weight loss. On the other hand, Diet A is the only diet that has less mean weight loss than the overall mean weight loss.

Question #4: Perform a one-way ANOVA of WeightLoss with the type of diet as the grouping variable. Check first if assumptions are met. Interpret the result.

```
# Test for Normality
WEIGHTLOSS %>% group_by(Diet) %>%
  summarize(shapiro_test(WeightLoss))
```

```
## # A tibble: 3 x 4
##   Diet  variable   statistic p.value
##   <fct> <chr>          <dbl>   <dbl>
## 1 A     WeightLoss     0.971 0.768
## 2 B     WeightLoss     0.971 0.781
## 3 C     WeightLoss     0.854 0.00626
```

The third group has a p-value < 0.05, violating the assumption of normality. We proceed with testing for homoscedasticity. We proceed to Levene's Test for Homoscedasticity

```
#Test for Homoscedasticity
levene_test( WEIGHTLOSS, WeightLoss ~ Diet)
```

```
## # A tibble: 1 x 4
##     df1   df2 statistic     p
##   <int> <int>     <dbl> <dbl>
## 1     2    57    0.0261 0.974
```

Since the p-value is > 0.05, we do not reject the null hypothesis. At 0.05 level of significance, we have sufficient evidence to conclude that the variances are equal.

We proceed with performing one-way ANOVA.

```
anova_diet <- aov(WeightLoss ~ Diet, WEIGHTLOSS)
summary(anova_diet)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Diet         2  57.45  28.727   9.124 0.000365 ***
## Residuals   57 179.47   3.149
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: Since the p-value is < 0.05, we reject the null hypothesis. At 0.05 level of significance, we have sufficient evidence to conclude that at least one of the means of the 3 diets is different from the rest.